



USAID
FROM THE AMERICAN PEOPLE



PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM

Semi-Annual Report, April 2014

30 APRIL 2014

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by NORC at the University of Chicago.

PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM SEMI-ANNUAL REPORT, APRIL 2014

30 April 2014

PN 7384; USAID Contract N0: AID-617-C-12-00006

PRESENTED TO:

USAID/Uganda

Joseph Mwangi

PRESENTED BY:

NORC at the University of Chicago

Jeffrey Telgarsky

Executive Vice President for Research

4350 East-West Highway, 8th Floor

Bethesda, MD 20814

Telephone: (301) 634-9413

Fax: (301) 634-9301

DISCLAIMER

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

HIGHLIGHTS	1
DESCRIPTION OF ACTIVITIES	2
A. Impact Evaluation Activities.....	2
Sampling for EGRA Cluster 2 Baseline Data Collection	2
Data Quality Assessment.....	2
Impact Analysis and Impact Evaluation Report.....	4
B. Performance Evaluation and CLA Activities	4
Systematic observation and documentation of project implementation activities as inputs for the performance evaluations.....	4
Collaboration, Learning and Adaptation (CLA) Activities	5
Performance Evaluation Planning	6
C. Risks to the Impact Evaluation.....	7
ANNEX 1. REVISED DATA COLLECTION PLAN (FROM SHRP'S SEPTEMBER 2013 PMP).....	10
Result 1 – Reading Program	10
Result 2 – School Health Program.....	11
ANNEX 2: FEEDBACK PRESENTED TO THE SHRP TEAM FOLLOWING OBSERVATION OF ENUMERATOR TRAINING AND PILOT TEST FOR CLUSTER 2 EGRA DATA COLLECTION	12
Positive feedback.....	12
1. EGRA C2 Baseline Assessors and Supervisors Training	12
2. EGRA C2 Baseline Data Collection	12
Constructive feedback on areas that can be strengthened.....	13
1. EGRA C2 Baseline Assessors and Supervisors Training	13
2. EGRA C2 Baseline Data Collection	14
Other	15
EGRA administration of certain subtasks	15
ANNEX 3: CONTENT OF MEMO TO USAID ON RISKS TO IMPACT EVALUATION RESULTING FROM EGRA IMPLEMENTATION PRACTICES OBSERVED DURING ENUMERATOR TRAINING, PILOT TEST, AND DATA COLLECTION.....	16
ANNEX 4: THREE REVISED OBSERVATION TOOLS.....	24
(1) Training Observation Tool: EGRA - R1	25
(2) Training Observation Tool: hiv/aids – R2	29
(3) Data Collection Observation Tool: EGRA and KAP.....	34

ANNEX 5. ANNEX 5: REVISED TEMPLATE FOR MONTHLY REPORTS PREPARED BY LOCAL STAFF AS INPUTS TO FEEDBACK MEMOS	36
ANNEX 6: CHALLENGES TO THE IMPACT EVALUATION, AS PRESENTED IN SEMI-ANNUAL REPORTS PRESENTED IN JUNE 2013 AND OCTOBER 2013	37

NORC at the University of Chicago (NORC), in partnership with the Panagora Group, is pleased to submit to USAID/Uganda this Semi-Annual Report recording progress made on the P&IE project between November 1, 2013 and April 30, 2014.

HIGHLIGHTS

During this reporting period, the NORC/Panagora team:

- Completed a data quality review of the EGRA Cluster 1 Round 2 dataset. This data, which will be used, along with Cluster 1 baseline data, to measure the impact of SHRP's Reading (Result 1) interventions, was collected by the IP in October-November 2013. NORC received the dataset in January 2014, and reviewed consistency across the baseline and Round 2 datasets in preparation for the impact analysis.
- Completed the first annual impact evaluation analysis for Result 1 using the baseline and Round 2 data from Cluster 1 schools collected in February-March and October-November 2013, respectively. Prepared and submitted to USAID the first Impact Evaluation Report for the P&IE Project, documenting progress in outcome indicators during the first year of SHRP implementation. Since only baseline data is available for the School Health (Result 2) at the time of this report, the impact analysis for Result 2 will be presented in the third annual impact evaluation report after follow-up data has been collected in October 2015.
- Provided the SHRP M&E team with support on sampling for the Cluster 2 baseline surveys
- Conducted numerous data quality assessment (DQA) tasks related to the Cluster 2 EGRA for which data collection took place in February 2014:
 - ▶ As mentioned above, conducted a data quality review of the Cluster 1 baseline EGRA and KAP datasets. Overall, we found the data to be of high quality.
 - ▶ Reviewed and provided feedback to the Implementing Partner (IP) on data collection instruments for the Cluster 2 baseline EGRA data collection, both from the perspective of NORC's role as evaluator and data quality reviewer.
 - ▶ Reviewed enumerator training manuals for the Cluster 2 baseline data collection.
 - ▶ Participated in enumerator training for the Cluster 2 baseline data collection, which took place in February 2014.
 - ▶ Travelled to a selection of districts in various regions to observe data collection activities for the EGRA data collections.
 - ▶ Identified and brought to the attention of the IP and USAID, several issues with EGRA data collection and sampling that could pose threats to the impact evaluation.
- Further refined the process of third party monitoring and performance feedback. For example, we have modified all our observation and monitoring tools to ensure that the purpose, context, methodology, and relationship to the work plan and PMP are clear to the reader. Our third party monitoring process is now a well-established routine of meetings and actions, based on a monthly calendar and an events and assignments tracker to ensure comprehensive monitoring.

- Continued to observe RTI/SHRP implementation activities, including training of trainers, training of teachers, EGRA data collection, and a stakeholders meeting. During this period, ten events were observed.
- Continued to record, using our observation tools, information and observations on each activity for the mid-term and final SHRP performance evaluation; and noted appreciative and constructive comments to provide as monthly performance feedback to RTI within the context of the CLA process.
- Continued to conduct monthly performance feedback meetings with SHRP Chief of Party and M&E Director, as well as key Results 1 and 2 team members based on observations from SHRP events and activities. During the reporting period we held four feedback meetings with RTI/SHRP.
- Developed the Mid-Term SHRP Performance Evaluation Implementation Plan.
- At the request of USAID, prepared a PowerPoint presentation on preliminary impact results of the SHRP Result 1 interventions, which was presented by USAID/Uganda at the Africa Regional Education Workshop in Accra, Ghana in March 2014.

DESCRIPTION OF ACTIVITIES

A. IMPACT EVALUATION ACTIVITIES

Sampling for EGRA Cluster 2 Baseline Data Collection

The P&IE Evaluation Expert continued to support the SHRP M&E Team Lead on sampling questions related to the Result 1, Cluster 2 Baseline. As described in the October 2013 Semi-Annual Report, after abandoning the multi-arm treatment approach for Cluster 1 schools, the IP intended to have 3 different treatment arms, with 56 schools per treatment arm (a total of 168 treatment schools), for Cluster 2. However, prior to data collection, the IP decided to limit intervention in Cluster 2 schools to a single arm treatment but to collect enough data to facilitate analyses at the district level given that the MoES has expressed interest in district-level results; hence NORC conducted new sample size calculations to facilitate this new requirement. The results indicated that the target sample size should include 14 treatment and 14 control schools in each of the 10 districts in Cluster 2. Due to various constraints that SHRP faced, the final target sample ended up being 272 schools rather than the 280 originally calculated. As indicated in previous semi-annual reports, for Cluster 2 and beyond, the IP will no longer collect data in comparison districts. Hence, the scope of the evaluation findings will be reduced after Cluster 1.

Please see Annex 1 for current data collection plans for both EGRA and KAP surveys.

Data Quality Assessment

During this reporting period, P&IE staff engaged in various data quality assessment (DQA) tasks related to the Cluster 1 EGRA datasets and Cluster 2 data collection. They included the following activities:

- Conducted a data quality review of the Cluster 1, Round 2 EGRA dataset. We also reviewed the Cluster 1 baseline and Round 2 datasets for consistency. NORC staff noted a few consistency and quality issues, but overall, we found the data to be of high quality. NORC staff took the data quality issues into account when providing advice and guidance during the training and pilot test for the EGRA Cluster 2 data collection.

- Reviewed and provided feedback on all data collection instruments (EGRA and learner context instruments, teacher/head teacher survey, classroom observation tool, school inventory) and training manuals for the Result 1, Cluster 2 EGRA data collection and provided written feedback to the SHRP team in January 2014 prior to the enumerators training. In providing feedback on instruments we were cognizant of maintaining consistency in instruments across clusters. We also noted that our comments to the training manuals had been made on previous iterations of the document, but had not been incorporated into the manuals as requested. The IP assured us that these changes would be made in time for the Cluster 2 baseline training. Additionally, a number of NORC's comments on the data collection instruments were taken into account in the final version of the questionnaires fielded for the Cluster 2 baseline.
- Participated in the enumerator training for the Cluster 2 baseline data collection. The P&IE team's Evaluation Expert and Survey Expert travelled to Uganda on February 1-14 to participate in the eight-day enumerator training and pilot test for the cluster 2 baseline EGRA data collection. NORC staff observed many positive aspects to the training and pilot tests, as well as several quality issues that could potentially have a detrimental effect on the quality of data collected. These issues were brought to the attention of the SHRP team, and some were addressed during the training, as feasible. Both positive and constructive feedback provided to the IP following the training and pilot test are presented in Annex 2.

Of particular concern were several issues related to guidelines given to enumerators about the implementation of certain EGRA subtasks, which could potentially have negative implications for the rigor of the impact evaluation. P&IE staff laid out these issues in a memo submitted to USAID on February 21, 2014. (Please see Annex 3 for the content of this memo).

- Conducted field observations in Manafwa district on February 19 – 20. NORC's Resident Evaluation Manager travelled to the field to observe field work for the EGRA data collection. Observations resulting from these field visits are presented in Annex 3.

In addition to the issues listed in Annex 3, we also observed an incident in one school, where there were some disruptions in the classroom as a result of interviewed students returning to class with pencils and/or biscuits, which they receive as thank-you tokens. The incident raised our concern for protection of human subjects in the course of the data collection. We raised the possible safety concerns for children associated with these interview incentives with the SHRP COP and M&E Team Leader, and received assurance that these incidents were isolated and not a cause for concern. USAID is aware of the reported incidents.

The incident also led us to understand that we were not fully aware of the IP's IRB-approved protocols. We received the IP's IRB submission in March, 2014 and noted several issues that might be out of compliance. As we do not have oversight of this element of the IP's activities, we provided detail of our concerns to USAID.

Impact Analysis and Impact Evaluation Report

P&IE staff conducted extensive analysis of the baseline and Round 2 EGRA data from Cluster 1 schools collected in February and October 2013 for the first annual impact evaluation of Result 1 activities conducted under SHRP. This included data cleaning and preparation of datasets for analysis, and subsequent analysis consisting of ordinary least squares multinomial regressions to evaluate the impact of SHRP on early reading skills. Different models were used to check the robustness of the results. The approach follows very closely the original analysis plans outlined in the approved evaluation design. We prepared and submitted to USAID the first Impact Evaluation Report for the P&IE Project, documenting the impact of SHRP in outcome indicators during the first year of implementation. Since only baseline data is available for Result 2, the impact analysis work on this data will be presented in the third annual impact evaluation report after follow-up data is collected in October 2015.

B. PERFORMANCE EVALUATION AND CLA ACTIVITIES

Systematic observation and documentation of project implementation activities as inputs for the performance evaluations.

The P&IE team, led by subcontractor Panagora Group, continued to implement our processes for systematically monitoring and documenting SHRP implementation activities, and our system for carrying out the CLA elements of our contract and providing RTI with performance feedback.

During this period, we continued to follow a sequence of information collection, reporting, review, and feedback that includes the following:

- A monthly meeting between in-country P&IE staff (Resident Evaluation Manager and Sr. HIV/AIDS Specialist) to review the prior month's work and determine content of the monthly report and the performance feedback memo to RTI based on information from the observation reports
- Preparation and submission of a monthly activities report and draft SHRP performance feedback memo by P&IE country staff to U.S.-based

Box 1: Tools and reports used by the P&IE team for observing & documenting implementation activities

SHRP Events and Assignment Tracker. We continued to maintain and update a SHRP Events and Assignment Tracker which serves as a management tool to ensure coverage of events by our Resident Evaluation Specialist and Senior HIV/AIDS Evaluator and to track submission of all observation reports. The tracker lists the name of the event, date, a description and type of event, duration, observation tool used, the observer assigned to the event, date report submitted by our in-country staff, and if applicable whether an SHRP training report was obtained from RTI. SHRP Events and Assignment Tracker continued to be updated and utilized for forward planning during our monthly P&IE team meetings. During this period the team continued to experience some delays and challenges in receiving upcoming SHRP planned activities.

Observation tools. During this period, the P&IE team refined the observation and monitoring tools used by our Resident Evaluation Specialist and Senior HIV/AIDS Evaluator, specifically to ensure that an overview and summary of observations is provided; and that the specific purpose, content, methodology, and relationship of the observed activity to the work plan and PMP are clearly indicated. These modifications have greatly helped readers understand better the reports of observed events. Specifically, we refined the following tools appended in Annex 4: Training Observation Tool, EGRA - R1; Training Observation Tool, HIV/AIDS - R2; and Data Collection Observation Tool, EGRA and KAP.

P&IE document binder. To facilitate access and review of implementation progress, in particular in anticipation of the imminent mid-term performance evaluations, all reports generated continue to be systematically organized and numbered as soft copies and hard copies into a binder.

Monthly report. Our in-country staff prepares monthly reports, in which they review the previous month's SHRP activities, and generates a first draft of the feedback memo to RTI. During this period we refined the monthly report template used by our in-country staff to ensure the level of detail needed (see refined template in Annex 5). The monthly reporting improved as a result.

team

- Full P&IE team meeting to discuss performance feedback memo and monthly in-country activities (U.S.-based and in-country staff)
- Monthly performance feedback meeting with SHRP project staff
- Performance Evaluation/CLA coordination meeting (Panagora, in-country staff)

The memos, meetings and feedback activities listed above have the end goal of accurately documenting the P&IE team's observations of implementation activities and providing appreciative and constructive feedback to the SHRP team, so they can use our observations and suggestions to improve implementation in real time. The memos have also proved to be invaluable in preparing for the Midterm Performance Evaluation to be conducted in June 2014.

During this reporting period, our in-country staff attended and observed the following meetings, events, and trainings, and prepared a report on each of them using the appropriate observation and monitoring tool. Each of the reports provided appreciative and constructive observations that were collated and shared with RTI as performance feedback.

- Cluster 1 Language Board Meeting that was organized in Masindi district for the Runyoro-Rutooro region on 11/11/2013
- Early Grade Literacy Master training for Cluster 1 (P1 additional schools & P2) and Cluster 2 (P1) materials on 12/16- 12/17/2013
- SHRP Implementing Partner Meeting on 12/10/2013
- PIASCY training for primary and post primary school teachers 1/14/2014- 1/15/2014
- Early Grade Literacy Training of Trainers (TOT) on C1 (P1 & P2) and C2 (P1) SHRP books; Shimoni PTC sessions 1/22/2014-1/23/2014
- Early Grade Literacy Training of Teachers on C1 (P1 & P2) and C2 (P1) SHRP books observed Nyondo PTC on 28 January; and Ngora PTC on 1/29/2014
- Orientation of district administrators and Core PTC officials during the training of Teachers at Regional Level 1/29/2014
- EGRA C2 Baseline Assessors and Supervisors training on 2/3/2014– 2/12/2014
- Meeting of MoES HIV Technical Working Group on 2/18/2014
- EGRA C2 Baseline Data Collection in Manafwa district on 2/19/2014-2/ 20/2014

Collaboration, Learning and Adaptation (CLA) Activities

To implement the CLA component of the P&IE contract, we continued to provide performance feedback to RTI on a monthly basis, with both appreciative and constructive feedback, focusing on elements of performance where real-time feedback will help to strengthen performance and lead to optimal outcomes.

The performance feedback continues to be drawn directly from the reports of meetings, events, and activities observed by P&IE in-country staff. Each observation tool includes a section to note and record both appreciative and constructive feedback. Our Resident Evaluation Manager and Senior HIV/AIDS

Evaluator continued to collate these comments into a monthly performance feedback memo, which is reviewed and finalized taking into consideration comments from the full P&IE team. This memo is then provided in advance to the SHRP team, as the basis for feedback sessions with project leadership and the full P&IE team participating.

During this period we refined the feedback memo template to include, for each meeting, event, or training in which feedback is provided, a description of the purpose, content, and methodology used, as well as the relationship of the meeting, event, or training to the work plan. Following the session, a summary of the discussion is inserted into the performance memo.

Four feedback sessions were held during the reporting period. These sessions prove to be mutually useful for both SHRP and P&IE staff; they provide an opportunity for SHRP staff to gain third party insights into implementation issues that allow them to improve their performance in real time; and for P&IE staff to obtain useful clarifications on implementation processes.

Performance Evaluation Planning

During this reporting period, the PE Team, led by Panagora Group, prepared a detailed “Mid-Term Performance Evaluation Implementation Plan” for SHRP. This plan is presented separately as a P&IE contract deliverable.

We coordinated with all stakeholders in setting the exact timing of the performance evaluation, taking into account USAID/Uganda’s preferences as well as the school calendar, national holidays, and availability of USAID and SHRP personnel. Through the various activities carried out under P&IE, in particular the third party monitoring and performance feedback, the P&IE team has acquired a high level of familiarity with SHRP and achieved an effective and well received way to provide performance feedback to its senior staff, while maintaining objectivity and independence. As described above, in the course of our P&IE work, we continuously review implementation documents and performance and therefore are able to carry out the Mid-Term Evaluation in an efficient manner, per a detailed timeline provided in the Mid-Term Performance Implementation Plan. The bulk of the work will be carried out from mid-June through early July 2014 in Uganda, including data collection at the national, district, and school levels.

The Mid-Term Performance Evaluation of SHRP will assess the effectiveness of activities to date, both from a standpoint of implementation and achievements, using both qualitative and quantitative data and identify progress in achieving its planned five-year results against SHRP’s planned achievements for this stage of the work. As a mid-term evaluation, it will also provide recommendations to maintain or improve implementation processes and progress in achieving results. Our Performance Evaluation team will carry out a cross-sectional descriptive and analytical performance evaluation employing qualitative methods of data collection supplemented with quantitative data gathered through review of documents. We will also examine the project’s inputs, processes and outputs. The data collection methodologies to be utilized include; a) document review; b) key informant interviews (KIIs); c) classroom observation; and d) focus group discussions (FGDs). The Mid-Term Performance Evaluation Implementation Plan describes the P&IE team’s approach in detail, and includes draft data collection instruments for key informant interviews, focus group discussions, and classroom observations, and an initial list of contacts. It also sets out key questions for the evaluation, namely:

- *Design.* Does the project design and structure adequately support and facilitate achievement of the desired results?
- *Implementation.* Is SHRP being implemented according to plan? Is the program on track to achieving its overall objectives and results? Are the systems and mechanism for documenting lessons learned and best practices effective?

- *Impact.* What are the key factors for differences in performance (reading skills acquisition and HIV/AIDS knowledge and skills) in schools receiving the same interventions?
- *Sustainability.* To what extent will the programs components and subcomponents continue without USAID assistance?
- *Cost-effectiveness.* What are the implications and recommendations for potential scale-up of program interventions? In what ways can the programs be more cost effective?
- *Management/Coordination/Lessons Learned.* How can program design, management and execution become more efficient toward achieving program goals?

Three P&IE team members, Team Leader/Performance Evaluation, Betsy Bassan, Literacy/Education Evaluation Specialist, Brenda Sinclair, and Senior HIV/AIDS Evaluator, Stella Neema will conduct the Mid-Term Performance Evaluation. All three have been associated with P&IE from the outset. In addition, the P&IE COP, Alicia Menendez, and NORC's home office P&IE director, Varuni Dayaratna, will provide ongoing oversight and support. Country-based support will be provided by NORC's Resident Evaluation Manager, Evelyn Namubiru, who will also participate in district and school site visits to expand the number of schools visited.

C. RISKS TO THE IMPACT EVALUATION

Below we present an assessment of risks/challenges to the impact evaluation design that emerged during this reporting period. Challenges identified prior to the current reporting period, which were presented in previous semi-annual reports, are listed in Annex 6.

- I. Result 1: During the Cluster 2 EGRA training and pilot test, the P&IE team observers noted some issues related to the implementation of three specific EGRA subtasks – Letter Sound Knowledge and Word Segmenting, and Oral Passage Reading - and that could have negative implications for the impact evaluation. Annex 3 describes the issues in great detail and also lays out the implications for the impact evaluation. In short, SHRP was using very stringent requirements for accepting letter sounds as correct; for example, while the EGRA toolkit states that “For consonants that can represent more than one sound (i.e., c, g), either answer is acceptable. For vowels, either the short or long sound is accepted (/i/ as in pin or as in pine),” in the SHRP implementation of EGRA only one sound per vowel was being accepted as correct. As well, local pronunciations of words – e.g. “muzzah” for mother – were being marked as incorrect. This raises the concern that learners who actually know correct letter sounds are assessed as not knowing them, since trainers were instructed during training to mark as wrong any very slight deviation from the “ideal” sound of a letter.

This approach can bias the assessment in favor of treatment schools, where students are being taught one correct letter sound or a specific pronunciation of a word, relative to control schools, where a broader set of letter sounds and pronunciations are being taught. We can take as an example the letter B¹: the sound of letter B is /b/ or /buh/². Both sounds are correct and

¹ Similar problems exist with many other consonants such as D, T, P, K, G, etc.

² RTI International, EGRA Toolkit, March 2009
<https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149>

accepted as building skills towards early reading ability. However, the current application of EGRA in Uganda only accepts a perfect clipped sound /b/ as correct. Marking /buh/ as wrong is likely to punish learners in control schools more than it punishes learners in treatment schools, because teachers in treatment schools are trained to teach /b/ as the only correct sound while teachers in control schools are likely to use either /b/ or /buh/ given that both sounds are considered correct. This approach of “teaching to the test” will bias impact findings in favor of treatment schools. NORC is exploring options for measuring this bias in order to adjust impact measures; towards this end, we briefly discussed some alternatives with USAID, such as measuring the bias by conducting experiments to test more and less restrictive versions of EGRA administration.

2. **Result 1: Possible contamination of controls.** Because the SHRP team is not planning to expand SHRP implementation to additional districts for Cluster 1, they are planning to implement Result 1 activities in control CCTs in the 11 original districts starting in 2014 in order to meet target numbers of trained teachers. However, they plan to exclude the control schools within the control CCTs which were selected for the EGRA data collection and intervene only in the schools from control CCTs which have not been included in the EGRA data collection. Hence, according to the SHRP M&E Team Lead, no teachers in any grade (P1 through P4) in the EGRA control schools will be trained; nor will instructional materials be distributed to these schools. CCTs associated with these control clusters will be strictly instructed not to provide any assistance to these control schools.

Strict exclusion of control schools from treatment is critical for the integrity of the impact evaluation design. While SHRP staff has assured us that no control schools will receive any semblance of the Result 1 interventions, we are nonetheless concerned by the possibility of contamination through CCTs or spillover of materials. Any contamination of the control schools will lead to underestimation of the effects of the SHRP Result 1 interventions. We have made this concern clear to both the IP and USAID, and requested that SHRP put in place adequate safeguards to ensure that the control schools in our sample will not be contaminated.

3. **Result 1: Non-systematic replacement of sample schools.** During Cluster 2 baseline data collection in Mbale district, the SHRP team opted to exclude control schools that use or were presumed to use Luganda and English instead of Lumasaaba as the medium of instruction. The appropriate procedure to replace these schools (following the replacement rule provided) was not followed. Two of these non-Lumasaaba instruction schools were replaced by schools in which the medium of instruction is Lumasaaba; these replacements were picked from the list of preselected schools designated as replacements. The rest of the non-Lumasaaba instruction schools in the district sample were neither assessed nor replaced. We indicated to the IP and USAID that this approach was neither appropriate to keeping the integrity of a random sample nor conducive to comparing SHRP schools to the average public school in Uganda. First, replacing sample schools with hand-picked replacements creates problems with the sample balance. Second, the aim of the evaluation is to assess reading ability of learners in English and local language. While it is not possible to test them in the local language (Lumasaaba, in this case) in schools that do not teach in Lumasaaba, it would still have been possible to test student's performance in English. As such, NORC's Evaluation Expert urged SHRP staff to conduct the EGRA in English in these schools as soon as we learned of the situation. However, the SHRP team did not comply with this request in a timely manner. Therefore, NORC decided that the impact analysis will need to exclude Mbale district altogether.

4. Result 1: Manafwa district is encountering a serious crisis created by teacher transfers in the region. We learned during field observations that most of the teachers trained by SHRP in January 2014 in this region have been transferred to other schools: four of the treatment schools visited by our local staff did not have a trained PI teacher, because s/he had been transferred. It will be critical to have information about the whereabouts of teachers trained by SHRP, since transfers of trained teachers away from treatment schools will have a severe effect on the impact evaluation. If these teachers end up at control schools, the impacts will be even more skewed. We will work with the IP and through our performance evaluation to try and capture the movement of trained teachers between schools.

ANNEX I. REVISED DATA COLLECTION PLAN (FROM SHRP'S SEPTEMBER 2013 PMP)

RESULT I – READING PROGRAM

Adjusted Early Grade Reading Assessment Data Collection Plan: 2013-2016

This table only includes data collection that is relevant to the impact evaluation.

	2013		2014		2015		2016	
	FEB	OCT	FEB	OCT	FEB	OCT	FEB	OCT
Cluster 1 A (4 languages)								
Treatment	P1:30	P1:30		P2:30		P3:30		P4:30
Control w/in district	P1:30	P1:30		P2:30		P3:30		P4:30
Control out district	P1:30	P1:30		P2:30		P3:30		P4:30
# of schools	280	168		TBD		TBD		TBD
Cluster 2 * (4 LANGUAGES)								
Treatment			P1:30	P1:30		P2:30		P3:30
Control w/in district			P1:30	P1:30		P2:30		P3:30
Control out district			----	----		----		----
# of schools			TBD	TBD		TBD		TBD
Cluster 3 (4 LANGUAGES)								
Treatment					P1:30	P1:30		P2:30
Control w/in district					P1:30	P1:30		P2:30
Control out district					----	----		----
# of schools					TBD	TBD		TBD

* As noted before, the IP no longer intends to collect data in 2016 for Cluster 2.

RESULT 2 – SCHOOL HEALTH PROGRAM

HIV and AIDS Assessment Data Collection Plan: 2013-2016

	2013		2015	2016
	BASELINE		MIDLINE	ENDLINE
	JUNE	OCT	OCT	OCT
Cluster 1				
Treatment	P4-P7 S1-S5 (partial)	S1-S5 (supplement)	P6-P7 S1-S5	P7 S1-S5
Control w/in district				
Control out district				
# of schools	232 primary + 232 post-primary		TBD	TBD
Cluster 2				
Treatment		P4-P7	P5-P7	P6-P7
Control w/in district				
Control out district				
# of schools		69 primary	TBD	TBD
Cluster 3				
In reviewing the data collection timeline in the September 2013 PMP, we note that the Result 2 intervention no longer occurs in the Cluster 3 districts/schools. This is a significant change in the implementation plans.				

ANNEX 2: FEEDBACK PRESENTED TO THE SHRP TEAM FOLLOWING OBSERVATION OF ENUMERATOR TRAINING AND PILOT TEST FOR CLUSTER 2 EGRA DATA COLLECTION

POSITIVE FEEDBACK

1. EGRA C2 Baseline Assessors and Supervisors Training

- Trainers and assessors alike were very eager and energetic throughout the training period. Assessors worked hard to learn the skills they were presented and trainers spent time with groups and individuals to practice skills throughout the training period. Trainers clearly had experience with the letter-sound and other sections of the assessment instrument, owing to their roles as the teacher trainers and to their experience training assessors for the Cluster 1 data collections in 2013.
- The pilot activity was very instructive for the assessors to learn how the interaction with the pupils would take place in the field. Assessors had an opportunity to learn the difficulties of ensuring that pupils understand and can react to the assessment, and they experienced the issues of students being unable to answer any items.
- The IRR tests were also very useful, and the scripted IRRs at the end of the training were particularly well done. The results were presented the day after, which kept the issues fresh in the minds of the assessors. This was a particularly useful element of the training.
- The two training venues were very pleasant facilities, adequate in size and had all the necessary amenities to make the training comfortable for all.
- The materials were nicely packaged by language group and available to all trainees. The tablets were pre-loaded with a near-final version of Tangerine and functioned well. Adequate technical staff were on hand to reboot or troubleshoot when any problems arose (which were few). Slides were visible from the back of room and sound was loud enough through microphone for all to hear. Trainers and assessors were on time and ready to start after breaks.

2. EGRA C2 Baseline Data Collection

- Supervisors did well with introductions when they arrived at schools. They followed the survey introductory process, in absence of a school administrator; they could always utilize the teachers present to prepare for the interviews. Learners' interviews began immediately after the team sought clearance.
- We observed good team-work between supervisors and assessors. Assessors supported the supervisor with the sampling of learners, organizing of interview areas, and they were supportive in carrying out some learners' interviews.
- The assessors were conversant using the tablet to conduct interviews. They could easily log in and scroll through the subtasks, knew when to stop the learner and saved the data collected.
- The assessors were aware of how hard it was to begin a learner's interview without establishing a good rapport. They started rapport building right from time they picked a learner from class.

Some went forward to comfort the learner after providing a seat and before reading the learners agreement/consent.

- Assessors were knowledgeable about the right stimuli to provide to a learner during the different subtasks and provided it at the right time.
- The assessors were mindful to aid quiet learners move on with the subtasks. They would help pointing at next sound or word and requested them to move on in a familiar language.
- Learners interviews were mainly conducted outside, assessors tried to find quiet places far away from classes and the seating arrangement allowed good eye contact between assessor and learner.

CONSTRUCTIVE FEEDBACK ON AREAS THAT CAN BE STRENGTHENED

I. EGRA C2 Baseline Assessors and Supervisors Training

Preparation: The training of trainers was very short, primarily consisting of a review of the agenda for first two days of training and watching video of letter sounds was shown. Preparation did not include slides or role-playing. Consequently, there were issues that came up in the training (e.g., transitions between tasks, focus of training modules, etc.) which could have been avoided with adequate training of trainers. Examples:

- Assessors were invited to select their own language instrument instead of having a single standard instrument for the first run-through of the tablet version. This caused a lot of confusion because local language versions have a different numbering system than the English version.
- The training agenda indicated that supervisor training and ‘what happens when we arrive at the school’ would happen on the last day, but these were dropped due to lack of time. There needs to be at least one full day of training of trainers to allow a complete run-through of slides and role-plays (with all role-plays scripted).

Materials:

- There were errors or missing instructions on the paper version of the instruments and some inconsistencies with the tablet versions. Errors of this type have the potential to reduce the confidence of assessors in the instruments, so we would recommend allowing adequate time for a careful review before the next round of data collection, taking into consideration the large number of instruments that require review.
- The tablet version had a few errors that were corrected during training (in addition to language updating).
- There were no scripts for the demonstrations until the final IRRs. This resulted in some unrealistic and confusing demonstrations, particularly when volunteer assessors carried out the role-play demos.
- There was no mention of the field manual until the supervisor training on the last day. Assessors were not provided with the field manual during training. The supervisor training revealed that the generically-produced field manual contained many instructions that were not pertinent for the Uganda version of EGRA. Throughout the training, content-specific questions were answered by whoever had the microphone (with some interventions from other trainers) and some of the responses did not seem accurate. We recommend that the interviewer manual be updated and include a “QxQ” (Question-by-Question) explanation of how each question is to be treated by the

interviewer. The manual should be provided to assessors before or at the beginning of training and be considered required reading. Creating and adhering to this manual would reduce conflicting answers provided during training.

- All local language teams had many comments on the translations of the instructions in the instruments and time was spent making corrections and provided many changes to the RTI/CRS team. These issues should be dealt with prior to the next training and be organized as a methodologically sound translating task, as finalizing the text is crucial for the assessment, and the assessors are not all language experts.
- Training videos were difficult to hear and understand. Both the sound-letter video and the videos of sampling were shown to assessors accompanied by some description. The sampling video was particularly difficult to understand and led to confusion on the part of assessors.

Activities:

- Trainers carried out role-plays to demonstrate the assessor-pupil interaction of the main instrument. Except for the final IRRs, these role-plays were not scripted, so they were a bit disorganized and the trainers did not introduce the tasks in advance of the demonstrations clearly. When volunteer assessors were demonstrating, they presented a number of errors. Demonstrating poor techniques before demonstrating good techniques tends to confuse interviewers, so we recommend trainers carry out scripted demos for the first three days and provide scripts to volunteer assessors.
- The consent/ introduction to the pupil instrument was reviewed quickly on the first day and skipped entirely when the assessors started using the tablets.
- There was some lack of organization for the pilot regarding which schools were selected. Cb confusion about which schools were to be used for the pilot led to some real sample schools being used for pilots, which required selecting replacement schools for the main data collection.

Supervisor training: The field manual made its first appearance at the supervisor training. Supervisors read from a few pages, although trainers quickly noticed that a number of the tasks listed for supervisors were only applicable to paper instruments or were only carried out by DQAs. The entire supervisor training lasted about an hour and there were few opportunities for questions from the new supervisors. We very strongly recommend that a minimum of 2/3 of a day of a well-organized training be dedicated to the supervisor training, as supervisor must guarantee the day to day data quality and logistics of their teams.

2. EGRA C2 Baseline Data Collection

- Some assessors were inconsistent in introducing and reading the instructions for the different subtasks as written on a tablet. Some dodged reading instructions and instead provided memorized instructions probably to shorten time spent reading. Others went through instructions very quickly. This confused some learners who therefore did not answer even after prompting. This in turn led some assessors to conclude that these learners were not able to complete “reading subtasks.” Inadequate introduction of subtasks can negatively affect learner performance.
- While administering the learners’ instrument, some assessors did not respect the administration rules (e.g., 3 and 5 seconds rules, when to tab the start button, and helping learners read). For example, we observed assessors who tabbed the start button before a learner said something; pushed learners to move through the reading tasks before 3 or 5 seconds elapsed; made learners

skip sounds/words in the process of aiding them to continue reading; or didn't note where a learner stopped.

- During the segmenting subtasks for both English and local language, some assessors failed to read the word twice as instructed, possibly to save time.
- Before the English vocabulary (subtask 5), some assessors would forget to place required objects (pencil with rubber and paper) in a visible place, making it harder for some learners to identify the objects. These learners were marked incorrect, even though they might have succeeded if the objects had been visible.
- In one school with a PI classroom lesson observation, we observed pencils and biscuits given to learners after attending an interview. This created a lot of distractions when the students returned to class, with some children having pencils and biscuits and others not. Given the setting, many learners come to school without morning tea or a packed snack, which significantly increases the value/desirability of the pencil and biscuit. We recommend that RTI find a way to provide such rewards so that they do not create disruptions back in the classroom (possibly giving them to all learners).
- DQA officers and other field managers should follow up on these issues. Standardization of EGRA administration, including reading instructions and allowing learners sufficient time to complete the tasks is a very important element of data quality. We would like to know what steps RTI has taken to ensure that these problems are resolved.

OTHER

EGRA administration of certain subtasks

We noted that certain guidelines given to enumerators about EGRA implementation raised concerns for the impact evaluation. These concerns mainly involve:

- The types of sounds accepted for the letter sound knowledge and segmenting subtasks
- The types of pronunciation accepted for words in the reading passage

We have provided detailed notes regarding the implication of these EGRA implementation guidelines on the impact evaluation in a separate memo that we shared with RTI and USAID on February 21st. NORC and RTI met to discuss the memo on March 18th.

ANNEX 3: CONTENT OF MEMO TO USAID ON RISKS TO IMPACT EVALUATION RESULTING FROM EGRA IMPLEMENTATION PRACTICES OBSERVED DURING ENUMERATOR TRAINING, PILOT TEST, AND DATA COLLECTION

Date: February 20, 2014

From: NORC at the University of Chicago

To: Joseph Mwangi, USAID

Cc: Saeeda Prew, RTI

Re: Issues with USAID/Uganda SHRP EGRA Implementation

I. Overview of the Issue

During the week of February 3-7 2014, the NORC team observed the EGRA assessor training facilitated by RTI. NORC noted some issues and is concerned about some aspects of the EGRA implementation. In what follows we describe the issues and their implications for the Impact Evaluation.

As a reminder, the EGRA tool was developed for English and each local language (LL). It is composed of several subtasks, namely: (1) letter sound knowledge, (2) word segmenting, (3) nonword decoding, (4) oral passage reading, (5) reading comprehension, (6) listening comprehension (only in the LL tool) and (7) receptive vocabulary (only in the English tool). In this memo, we focus on the English language EGRA tool and specific issues in the following subtasks:

(a) Letter Sound Knowledge and Word Segmenting

In the letter sound knowledge subtask, learners are presented with a set of letters and are asked to produce the sound associated with each letter. In the word segmenting subtask, learners are asked to break down words into distinctive sounds (phonemes). For both subtasks, in past EGRAs, more than one sound was accepted as correct for certain letters/phonemes; however in the EGRA implemented by RTI for USAID/Uganda, rules have become more restrictive such that only one specific sound for each letter is deemed correct.

(b) Oral Passage Reading

In the oral passage reading, learners are instructed to read a short story. Words are marked incorrect by enumerators if they are read incorrectly. In RTI's EGRA for

USAID/Uganda, enumerators have been instructed to also mark as incorrect words that are not pronounced as in Standard English. This is problematic because some learners have a strong accent (e.g. northern Ugandan learners) so they may not be able to pronounce the words with the pronunciation desired even if they can decode (and perhaps even comprehend) the words.

Below, we present the email exchange between NORC and RTI that details the issues outlined above, and discuss the implications for the Impact Evaluation.

II. Email Exchange between NORC and RTI

The following email exchange took place between NORC and RTI between February 7 and February 9, 2014. In these messages we explain NORC's concerns about letter sounds that are considered incorrect during the administration of the EGRA tool as well as other EGRA-related issues raised during training. We received one response from Peggy Dubeck, the RTI literacy expert, included below.

FEBRUARY 7

Hi Tracy,

This is the issue that concerns me about the EGRA training as it is being applied.

The EGRA in English includes a task which focuses on letter sound knowledge. According to the EGRA toolkit:

(<https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149>)

Item Construction. The same laminated page of letters used in the first subtest of letter name

knowledge should be used for assessing letter sound knowledge. For consonants that can represent more than one sound (i.e., c, g), either answer is acceptable. For vowels, either the short or long sound is accepted (lil as in pin or as in pine). Students may have difficulty in eliminating the vowel sound frequently associated with consonants; in these cases either /b/ or /buh/ is accepted as a correct response. During training, enumerators and supervisors should carefully review possible pronunciations of each letter. (For a complete listing of characters and symbols in phonetic alphabets, please see <http://www.antimoon.com/misc/phonchart2008.pdf>.)

However in Uganda EGRA C2 training different rules have been applied. Instead, only one sound per vowel is accepted. For example, the accepted sound for the letter 'i' is /ee/ as in INK but not "eye" as in the personal pronoun "I" or short i as in HIT. Similarly, the accepted sound for the letter 'u' is /ah/ as in UP but it is assessed as wrong if a pupil identifies the sound as /oo/

as in UGANDA (local pronunciation) or /ue/ as in UGANDA (British or American pronunciation).

Another example is letter Y which can be hard as it is in YELLOW or soft as in MY.

Assessors are also trained to assess as wrong any attempt to include even a small vowel sound after consonants such as b, p, d, etc.

Some of these rules have created confusion about the correct sound of some “non-words” that pupils are supposed to read aloud. For example, when reading the non-word TAW, should the pupil say /t/, /a/, /w/ using the sound /a/ as in APPLE or should she say /t/, /o/, /w/ similar to LOW? (which BTW it is how the word LAW is pronounced in Uganda)

This tool is used to follow the evolution of learners from P1 to P4. As children increase their ability to read they learn the variety of sounds that vowels and consonants make in English. The evaluation requires a consistent application of the rules year after year in order to be valid.

Please explain why the established rules for applying EGRA have not been presented during this training.

Thanks

Alicia

FEBRUARY 8

Dear Alicia, our RTI support team was able to rally and address the points you had raised. The below was written up by Peggy, but the sentiment from Ben is there also.

I am copying both Ben and Peggy on this strand so we can efficiently address any additional issues.

Thanks for your support.

Tracy

English letter sounds:

I. Why short vowel sounds only?

- a. If the child said the “long vowel”, it is conflated with the letter NAME. The long vowel sound of five English vowels is also their name. If we accepted the long vowel sound, a child could not know any letter sounds but respond with the letter names which would be then be score correctly for the task. And analysing the results would not be a meaningful.
- b. The short vowel sound is the sound of the letter when it is in isolation.

- c. English beginning reading includes primarily short vowel sounds (CVC consonant – vowel – consonants say the short vowel). The task is hoping to capture what beginning readers know. (and those in upper stages too)
2. Why only accept one sound per letter?
 - a. We only accept the hard g (as in gasp) and hard c (as in cat) because they are the most common sounds for those letters. They are taught first, and if we accepted both, that would confuse data analysis.
3. Why don't we accept other sounds for letters such as the long i in the word 'my'?
 - a. The answer can be explained by the task itself – we are testing the sound of individual letters. And the sound of long i in my is explained by its pattern (and a bit into meaning layer) . The word “my” is an open-syllable word. Many English syllables that are “open” (they don't end in a consonant) are pronounced long. (Other examples: because, me, go, table).
 - b. Yes, the letter 'y' in my says i but lots of English letters change their pronunciation based on the pattern of the word (example “to” says long u as in flute). Again, we accept the sound of the letter (the most common and its sound in isolation)
 - c. *If you want to know more: English spelling and pronunciation can be explained by a sound layer, a pattern layer and a meaning layer. Example spelling explained by the sound layer is “cat” – the three sounds in that word are represented by three letters. The pattern layer is the pattern (the order) explains how some words are pronounced and spelled. For example, when the letters “oa” are next to each other we get long o (boat) or when 'dge' are next to each other we get /j/ (bridge). Finally many words spelling or pronunciation is explained by a meaning layer. An is “confidence”. The letter “i” would be puzzling to some because it sounds more like a short u. But the “letter i” is used because the root word is confide.
4. Why do we want the clipped sound?
 - a. That is what we are teaching and we want to see intervention effects. Lydia can help to explain what can happen if children attach an extra vowel to each letter (the 3-syllable Luganda word bikopo could become a 6-syllable word baikaopao or buikuopuo).
 - b. Uttering some sounds (plosives p,b; dental t, d; velar k, g) will have some additional air partially because the mouth opens. But limiting what is considered correct responds to the instructional benefits mentioned in 4a and allows these results to be compared to previous collections.
5. Implied in this inquiry could be why not assess children's knowledge of letter names?
 - a. They could assessed alongside of letter sounds. To limit the amount of assessment time they have not been included in Uganda.
6. Why aren't we following the EGRA rules stated in the global toolkit developed in 2009?

- a. Because EGRA has evolved and improved. Since that toolkit was written, some procedures have changed to respond to increased ability by trainers to get assessors reliable, to modify procedures so they are more valid for the measuring the intended construct, or to respond to the context. The theoretical framework for EGRA was informed by several assessments – of reading, of phonology and, of vocabulary (informal IRI's, DIBELS, CTOPP, PPVT, DIBELS). Not all elements of those assessments were originally used. Some have been dropped, others originally not included added. Ultimately what matters is that we know what was assessed and the procedures the assessors used.
7. Finally to the need for the tasks to get more difficult as children progress from P1 to P4: One of the strengths of EGRA is that, by timing some of the tasks, we avoid ceiling effects that would be seen if they weren't timed while we are able to capture the phenomena that recognition of letters, words, and connected text becomes more automatic and accurate (i.e., elements of fluency) as reading skills advance by using the same items. In other words, a passage written at a grade 2 level might be read at 10 wpm in P1, 30 wpm in P2, and 60 wpm in P3. If we had unlimited resources and time, we might have a range of levelled stories from emergent (PrePrimer A) to advanced (grade 8) and we would find the highest level that a child can read with high accuracy 95%ish and report on two areas: 1) the instructional levels of the assessed children and 2) the percent of children who can read at grade level with 95%ish accuracy and high comprehension. Obviously, we don't have unlimited resources or time so we rely on using a timed passage and words in isolation as they give us good information and show growth. We may find that eventually, we want two passages for the P4 children, maybe written at staggered grade levels (2 and 4) to use with the kiddos who are over 80 wpm on grade 2 text.

FEBRUARY 9

Thank you very much for this information Peggy. I really appreciate yours and Ben's help with these issues. It is very helpful to learn that the EGRA has been modified and now it is implemented in a different way. Ben, is this also the approach you are using in Kenya?

I still have a few questions and comments.

- 1) In the light of revised accepted sounds for letters, how is the non-word TAW expected to be pronounced by the learner? and the non-word REW?
- 2) Demos and discussions of section 4a Oral Passage Reading focused on word pronunciation and fluency and included instructions to mark as wrong items that were poorly pronounced such as muzzah, fazzah, bruzzah. Some assessors indicated that those were common pronunciations of the words mother, father and brother in some areas/language groups. Should those pronunciations be marked wrong? Is reading the word with this pronunciation the same as not being able to read the word aloud at all? I think Tracy sent this question to you both last week.

3) Regarding your point 4a and 4b, I can see why you think it is best to teach students the sound of letter B as /b/ and train the teachers to use that approach in class. However, I am very concerned about using this approach to determine if a learner knows the letter's sound or not given that the sound /buh/ is also correct. Under this approach, the sound /buh/ would be marked incorrect as much as a completely different sound, a "don't know", or no answer.

We need an assessment tool that measures the learners' skills *regardless* of the method by which students are being taught. Please, remember that we are not just looking at the abilities of students in treatment schools or just checking if teachers follow SHRP instructions. We need to have an appropriate tool to conduct an impact evaluation of the program which can fairly evaluate pupils in treatment *and control* groups.

My concern is that limiting so much what is considered the correct sound of a letter can undermine the usefulness of the instrument as an assessment tool of the early reading skills of all learners. In training very strict limits of acceptability were enforced very strongly and any small deviation from the "ideal" sound of letters like b, g, h, t or p was marked as incorrect. This is likely to punish learners that know the letter sounds but pronounce them in a slightly stronger way.

I am much less concerned with the other letters such as a or i that I mentioned in my message. You are right that beginner readers would likely know the short vowels, so I think this won't be such a big issue.

I have been thinking about this long and carefully and that is why I didn't rush to respond. I am afraid that we can end up without a valid impact evaluation of the program which is USAID's requirement.

I know it is Sunday and I really appreciate everyone's efforts. If you think that it would be easier to use skype rather than email to discuss this, please let me know.

Many thanks!

Alicia

After this last email, we did not receive any response from RTI to the issues we raised. The words TAW and REW were substituted during the training presentation the next day by the words PEB and FID but we received no comments about the replacements and the methodology for choosing these replacements. We also do not know if these words were replaced in the pupil stimuli taken to the field.

III. Implications for the Impact Evaluation

To summarize, NORC is concerned about the possibility that learners that actually know correct letter sounds are assessed as not knowing them³. During training, any very slight

³ Please note that NORC is not expressing any opinion about how teachers and pupils should be trained on letter sounds. .

deviation from what trainers considered the ideal sound of a letter was instructed to be marked as wrong⁴ and this was very strictly enforced.

The reason behind our concern is that this approach can bias the assessment in favor of treatment schools and therefore produce a flawed evaluation. To make this clear we use an example, the letter B⁵. The sound of letter B is /b/ or /buh/⁶. Both sounds are correct and accepted as building skills towards early reading ability. However, the current application of EGRA in Uganda only accepts a perfect sound /b/ as correct. Marking /buh/ as wrong is likely to punish learners in control schools more than it punishes learners in treatment schools, because teachers in treatment schools are trained to teach /b/ as the only correct sound while teachers in control schools are likely to use either /b/ or /buh/ given that both sounds are considered correct.

The point 4a made by Peggy Dubeck in her email ("[That is what we are teaching and we want to see intervention effects](#)") explains why this strict interpretation of the correct sound is being used. Peggy is referring to checking if the learners in treatment schools are actually learning the sound that they are supposed to learn: /b/. This would be fine *if the goal of the evaluation were to confirm the program execution and achievement*. In that case the objective would only be to see if teacher training actually translates into pupils' knowledge of the selected sound /b/. In that case, there would not even be a need to include control schools in the evaluation.

But that is not the object of the impact evaluation. The purpose of the impact evaluation is to assess the effect of the program on early reading skills of pupils. In order to properly do so we need to compare pupils in treatment schools against pupils in control schools to see if the treatment schools are doing better than control schools in developing early reading skills among learners, and if so, how much better are they doing. The problem is that if both - /b/ and /buh/- are correct sounds for letter B and they both can be considered skills that translate into reading ability, then they both need to be accepted, otherwise the assessment itself biases the results towards the treatment schools and loses validity as an assessment tool to measure early reading ability *regardless* of the teaching method. An assessment that only accepts the sound intended to be used in treatment schools and nothing else, even if it is a correct sound, creates a bias. A biased assessment is not a valid instrument.

If there is a possibility of bias, any difference we observe between treatment and control will be "contaminated" and would not reflect the real difference between treatment and control schools. Without knowing how large the bias is, it is impossible to know the real effect the program makes. This clearly undermines any result and it is, of course, in no one's interest.

⁴ We are not referring to sounds such as /bah/ or /dah/ but instead, we mean slight vowel sounds after the consonant or slight sounds while trying to say the sound of letter H, for example. The preoccupation about getting the ideal letter sound was such that assessors were wondering if they should include in the assessment learners with missing front teeth because they probably were not able to make the /d/ or /t/ sounds correctly.

⁵ Similar problems exist with many other consonants such as D, T, P, K, G, etc.

⁶ RTI International, EGRA Toolkit, March 2009
<https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149>

In addition to the implications that the very limiting approach followed during training can have for the impact evaluation we are also concerned about the fact that it could also give the idea that learners -in general, either in treatment or in control schools- know fewer sounds than they actually do, misleading the interpretation of results and underestimating Uganda pupils' skills.

Finally, we also want to bring attention to what is considered a correct or incorrect pronunciation of certain words. As stated in the last email above, it seems that in certain population groups the pronunciation of the English word "mother" sounds something like "muzzah", where the TH sounds like a /zz/. Other examples are father or birth. In the oral reading passage, assessors were instructed to mark words wrong if they were pronounced with the /zz/ rather than /th/ sound, in the same way they should mark the work of a learner that is not able to read the word at all. We are concerned that this focus on pronunciation rather than oral reading ability will incorrectly suggest a child cannot read the word when, in fact, s/he merely pronounces it in the common manner. This interpretation of the assessment, like the letter sound issue we've described above, can bias the assessment if teachers in the program (treatment schools) are instructed to stop using the local pronunciation and to correct pupils when they do, while control schools will continue to consider "muzzah" an acceptable pronunciation when reading aloud the word "mother.". It can be an important bias given that the EGRA paragraph includes three words of this type: mother, father and brother.

Currently, NORC is exploring ways to still be able to produce a rigorous evaluation of the program

ANNEX 4: THREE REVISED OBSERVATION TOOLS

- (1) Training Observation Tool: EGRA - R1**
- (2) Training Observation Tool: HIV/AIDS - R2**
- (3) Data Collection Observation Tool: EGRA and KAP**

PROCESS EVALUATION
(I) TRAINING OBSERVATION TOOL: EGRA - RI

OVERVIEW

Purpose of training:

Relationship to work plan

Content (describe general content and include agenda where possible):

Methodology used for the training:

TRAINING DATA

Date of the observation: _____

Name of observer: _____

Training title/topic: _____

Name of facilitator 1: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Training title/topic: _____

Name of facilitator 2: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Training title/topic: _____

Name of facilitator 3: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Specific training objectives: _____

Number of participants: _____ (Male: _____ Female: _____)

Number of invitees: _____ Number of attendees: _____

Participants profile: Teacher: Primary _____ Post primary _____

School administrators (HT/DpHT) _____

Principals/College Tutors _____

CCTs _____

Language board member: _____



Other _____



Total number of project target districts: _____ Number of districts attending: _____

Total number of project target schools: _____ Number of schools attending: _____

OBSERVATION QUESTIONS

In order to evaluate the training workshop, check the boxes for the items you have observed. Use the “Answers and Notes” column to record interpretations and recommendations for future workshops.

Questions and considerations	Yes 	Needs more work 	Notes & Follow up (provide comments/feedback)
Did the facilitator(s) set-up the training workshop adequately (review objectives, expectations, ground rules, etc.)?			
Was there enough time allotted for each portion of the training?			
Was the training methodology appropriate? (describe the methods in the notes column)			
Were there sufficient resources (i.e., materials, aides, notebooks, flip charts) for all trainers and training participants?			
Was there lively interaction during plenary sessions? Did participants appear engaged in group work exercises? Did they ask a lot of questions?			
How well did the facilitator monitor the exercises?			
Was the debriefing done effectively amongst facilitators and RTI?			
Did the participants have an opportunity to practice skills?			
Was there a clear learning objective for each training session?			
Were the training objectives met?			
Were there follow-up actions/activities? What is expected after the training?			
Was there a pre-test and post test given to the participants?			

Questions and considerations	Yes 	Needs more work 	Notes & Follow up (provide comments/feedback)
Were results analyzed? (if yes, please provide results)			
Was the training monitored by the main coordinator RTI Result 1 and/or Result 2 staff? If so, provide names.			
Was the training monitored by the MOES staff?			
Was there anything that could be improved?			
Please provide any additional information on the observed session?			

Developed using information from: Levels of evaluation based on Kirkpatrick, D., 1994, *Evaluating Training Programs: The Four Levels*, San Francisco: Berrett-Koehler.

RATINGS OF KEY INDICATORS

	Not at all				To a great extent	Don't know	N/A
1. The strategies in this session were appropriate for accomplishing the training session's purposes.	1	2	3	4	5	6	7
2. The session effectively built on participants' knowledge of content, teaching, learning, and/or the reform/change process	1	2	3	4	5	6	7
3. Presenter displayed an understanding of pedagogical concepts (e.g., in his/her dialogue with participants)	1	2	3	4	5	6	7
4. The session's design encouraged a collaborative and participatory approach to learning	1	2	3	4	5	6	7
5. Participants appeared engaged in group work and plenary discussions	1	2	3	4	5	6	7

	Not at all				To a great extent	Don't know	N/A
6. The session's design provided opportunities for teachers to consider classroom application of resources, strategies, and techniques	1	2	3	4	5	6	7
7. Adequate time and structure were provided for reflection	1	2	3	4	5	6	7
8. Adequate time and structure were provided for participants to share experiences and insights	1	2	3	4	5	6	7

9. Overall achievement of training objectives 1 2 3 4 5

- Record any additional observations about the training process including general impressions; challenges encountered; dynamics among participants; hopes/fears expressed by participants; dominating and/or dull characters/presenters; etc.

AREAS TO RAISE WITH RTI

- Strengths:
- Areas of improvement:

ANNEX

Please append the following:

- Participants registration sheet (name, sex, designation, district, school, and telephone contact)
- Agenda

PROCESS EVALUATION
(2) TRAINING OBSERVATION TOOL: HIV/AIDS – R2

OVERVIEW

Purpose of training:

Relationship to work plan

Content (describe general content and include agenda where possible):

Methodology used for the training:

TRAINING DATA

Date of the observation: _____

Name of observer: _____

Training title/topic: _____

Name of facilitator 1: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Training title/topic: _____

Name of facilitator 2: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Training title/topic: _____

Name of facilitator 3: _____

Title: _____ Sex: _____

Affiliation/organization: _____

Specific training objectives: _____

Number of participants: _____ (Male: _____ Female: _____)

Number of invitees: _____ Number of attendees: _____



Participants profile: Teacher: Primary _____ Post primary _____
 School administrators (HT/DpHT) _____
 Principals/College Tutors _____
 CCTs _____
 Language board member: _____
 Other _____

Total number of project target districts: _____ Number of districts attending: _____



Total number of project target schools: _____ Number of schools attending: _____

OBSERVATION QUESTIONS

In order to evaluate the training workshop, check the boxes for the items you have observed. Use the “Answers and Notes” column to record interpretations and recommendations for future workshops.

Questions and considerations	Yes 	Needs more work 	Notes & Follow up (provide comments/feedback)
Did the facilitator(s) set-up the training workshop adequately (review objectives, expectations, ground rules, etc.)?			
Was there enough/too much time allotted for each portion of the training?			
Was the training methodology appropriate? (describe the methods in the notes column)			
Were there sufficient resources (i.e., materials, aides, notebooks, flip charts) for all trainers and training participants?			
Was there lively interaction during plenary sessions? Did participants appear engaged in group work exercises? Did they ask a lot of questions?			
How well did the facilitator monitor the exercises?			
Was the debriefing done effectively?			

PERFORMANCE & IMPACT EVALUATION (P&IE) SEMI-ANNUAL REPORT

Questions and considerations	Yes 	Needs more work 	Notes & Follow up (provide comments/feedback)
Did the participants have an opportunity to practice skills?			
Was there a clear learning objective for each training session?			
Were the training objectives met?			
Were there follow-up actions/activities? What is expected after the training?			
Was there a pre-test and post test given to the participants? Were results analyzed? (if yes, please provide results)			
Was the training monitored by the main coordinator RTI Result 1 and/or Result 2 staff? If so, provide names.			
Was the training monitored by the MOES staff?			
Was there anything that could have been improved?			
Please provide any additional information on the observed session?			

Developed using information from: Levels of evaluation based on Kirkpatrick, D., 1994, *Evaluating Training Programs: The Four Levels*, San Francisco: Berrett-Koehler.

RATINGS OF KEY INDICATORS

	Not at all				To a great extent	Don't know	N/A
1. The strategies in this session were appropriate for accomplishing the training session's purposes.	1	2	3	4	5	6	7
2. The session effectively built on participants' knowledge of content, teaching, learning, and/or the reform/change process	1	2	3	4	5	6	7
3. Presenter displayed an understanding of pedagogical concepts (e.g., in his/her dialogue with participants)	1	2	3	4	5	6	7
4. The session's design encouraged a collaborative and participatory approach to learning	1	2	3	4	5	6	7
5. Participants appeared engaged in group work and plenary discussions	1	2	3	4	5	6	7
6. The session's design provided opportunities for teachers to consider classroom application of resources, strategies, and techniques	1	2	3	4	5	6	7
7. Adequate time and structure were provided for reflection	1	2	3	4	5	6	7
8. Adequate time and structure were provided for participants to share experiences and insights	1	2	3	4	5	6	7
9. Achievement of training objectives	1	2	3	4	5		

- Record any additional observations about the training process, including general impressions and any challenges encountered, dynamics among participants, hopes/fears expressed by participants, dominating and/or dull characters/presenters, etc.

AREAS TO RAISE WITH RTI

- Strengths:
- Areas of improvement:

ANNEX

Participants registration sheet (name, sex, designation, district, school, and telephone contact)

PROCESS EVALUATION

(3) DATA COLLECTION OBSERVATION TOOL: EGRA AND KAP

Purpose of data collection:

Relationship to work plan:

Describe data collection event (how organized, conducted, and methodology used):

Observation date: _____

Data collection period: _____

District: _____

Local Language region: _____

Name of school: _____

Student population (by sex): _____

Students by Grade _____

Number of teachers in the school (by sex) _____

Name of Coordinating Centre: _____

Name of Sub-county: _____

GENERAL OVERVIEW OF OBSERVATIONS COLLECTED

Please provide a summary of your observations in this section.

OBSERVATION QUESTIONS

In order to evaluate the EGRA data collection, check the boxes for the items you have observed. Use the “Answers and Notes” column to record interpretations and recommendations for future workshops.

Questions & considerations	Yes	Needs improvement	Notes & follow-up (provide comments/feedback)
Did the school receive communication in advance on the data collection exercise (were administrators and targeted teachers aware beforehand)?			
Did the interviewers/supervisors comply with the survey sampling methodology (selection of pupils/teachers)?			
Was the interview environment sufficiently private (seating arrangement, room privacy, etc.)?			

Questions & considerations	Yes	Needs improvement	Notes & follow-up (provide comments/feedback)
Was rapport created with the interviewees (students/teachers/school administrators)?			
Other observations (e.g., school structures such as classrooms, desks, learning materials).			
What was the total number of pupils in attendance at the time of data collection? What is the regular total number of pupils for the observed class? (report by sex)			
Did the school administrators/teachers/school nurse attend the relevant SHRP training?			
Was there feedback provided from the assessors/supervisors? If so, how was feedback obtained? Please summarize the feedback, if provided.			
Please provide any additional information on the observed session.			

AREAS TO RAISE WITH RTI

- Strengths:
- Areas of improvement:

ANNEX

Please provide a hard copy of the tools/questionnaires used in data collection

ANNEX 5. ANNEX 5: REVISED TEMPLATE FOR MONTHLY REPORTS PREPARED BY LOCAL STAFF AS INPUTS TO FEEDBACK MEMOS

Submitted by, [Name]

(Note: please provide a brief report between 2-3 pages in length)

I. Time period covered:

II. Summary of key activities undertaken during this period (i.e. meetings, workshops, site visits, etc.):

For each activity observed during this period please include:

- Activity name
- Date observed
- Location
- Purpose of the activity
- Relationship to work plan
- Summary of observation comments (with note to see fuller report)

III. SHRP's overall progress against the work plan & PMP:

For this section please review SHRP's work plan and PMP and report whether or not SHRP was on track during the reporting period. Were planned activities delayed? Were activities implemented sooner than expected?

IV. Planned activities for the following period:

Activity	Result Area	Date	P&IE staff to observe the activity

V. Comments and concerns for discussion:

Please include (1) any concluding comments and concerns regarding previously observed events and/or (2) any issues or concerns regarding past or upcoming activities.

ANNEX 6: CHALLENGES TO THE IMPACT EVALUATION, AS PRESENTED IN SEMI-ANNUAL REPORTS PRESENTED IN JUNE 2013 AND OCTOBER 2013

1. **Result 1:** Delays in the implementation of Result 1 continued through this reporting period. Although all the trainings have taken place, including refresher TOT and teacher training on Cluster 1 P1 materials (teacher guides and primers), these instructional materials were still being distributed to Cluster 1 schools as late as September/October 2013. Our understanding is that materials have not reached all schools at the time of writing this report. In addition, the original plans that included three different treatment arms were modified and treatment was uniform across all schools. Baseline data collection for Cluster 1 was completed successfully in February, and follow-on data collection for Cluster 1 is being fielded among a sub-sample of primary schools. These delays and modifications to the implementation do not pose serious risks to the evaluation at this juncture. We plan to evaluate the impact of the program as it was implemented.

While the implementation changes/delays are not a risk to the evaluation design, an important fact to keep in mind, however, is that we do not expect to see the impacts of the full Result 1 intervention (teacher training and instructional materials) during this first impact analysis, using Oct/Nov 2013 data. However, the Oct/Nov 2013 data will provide us with an opportunity to measure the impact of multiple rounds of teacher training.

2. **Result 1:** The most recent version of the SHRP PMP indicates that no data will be collected from Cluster 2 in 2016. Going forward with this decision would imply that the impact evaluation for Cluster 2 would only be possible for P1 and P2 but not for P3. Given that Cluster 1 did not receive the full intervention in 2013, Cluster 2 will be the only group that will have a chance to receive three years of full treatment from the beginning of their primary education. The Evaluation Expert already mentioned this omission as a concern to USAID and to the IP as well.
3. **Result 1:** Data for the second EGRA wave are being collected as we write this report. Initial information from the field indicates low response rates (i.e. low numbers of students are being found) in the schools in the Central Region compared to baseline. We are currently working with the IP to try to address this problem and minimize the risks of having a small sample.
4. **Result 2:** There are several issues related to sample that have surfaced during the ongoing KAP data collection, which are likely to pose threats to the evaluation of Result 2 activities.

We noted in our first Semi-Annual Report that, it was not possible to include boarding or partial boarding schools -very common among post-primary establishments- in the evaluation sample, given delays in obtaining parental consent for the KAP Survey during the school year. We decided, however, to use the second round of the KAP survey (KAP2) to collect additional baseline data from Cluster 1 boarding and partial boarding post-primary schools by distributing parental consent forms to students before the school break. The idea was to ensure that the baseline survey consisted of a representative sample of post-primary schools, thereby allowing us to generalize the results of the impact evaluation to all such schools in the districts.

We recently learned of several problems that the IP is encountering with the supplemental boarding school component of the second round of KAP surveys. These problems could potentially have serious implications for sample size and the representativeness of the post-primary school sample:

- ▶ The IP faced resistance to data collection activities from some schools, where principals cited concerns that the survey would take away from exam preparation time (national exams in post-primary schools begin in the 2nd week of October) and some head teachers did not distribute consent forms to students at all. These schools could not be interviewed.
- ▶ Some schools closed before the end of the term and consent forms were not distributed on time. These schools could not be interviewed.
- ▶ Other programs related to HIV/AIDS have interacted with some of the schools and, therefore, head teachers decided not to participate in KAP. This is particularly the case of private secondary schools. These schools could not be interviewed.
- ▶ The sample frame that the IP provided NORC for selection of the school sample for the KAP2 contained errors; it included schools that already participated in the first round of KAP. In cases where these schools were randomly selected for the KAP2 sample, they had to be removed from the sample and, where possible, replaced.

NORC has requested from the IP a list of all schools in the KAP2 sample with disposition comments for each of the schools. After evaluating the situation we will have a clearer impression of the effect that these problems can have on the evaluation. At a minimum, we expect a reduction in sample size.

5. Result 2: As mentioned above, SHRP decided not to include post primary establishments in new treatment districts (Cluster 2 and after). Therefore, we will only be able to assess the impact of the Result 2 intervention on post-primary educational facilities for Cluster 1 schools.
6. Result 2: Based on the most recent PMP, we note that the Result 2 intervention will no longer be conducted in Cluster 3 districts and schools. As a result, NORC will focus its evaluation of Result 2 on Cluster 1 and Cluster 2 schools.
7. Result 1: Given program implementation delays in Year 1, the academic term was delayed for one week in the 11 districts of Cluster 1 where the IP is working in order to build in time to prepare and have teacher guides ready for the second training of teachers. Additional classes to compensate for the one week delay are not currently planned. An equivalent delay did not occur in the control district schools; therefore, the academic year in those schools will be one week longer. We do not anticipate a visible effect, but it is worth mentioning how the reality of the program may affect the evaluation.
8. Result 2: After NORC selected the samples for the impact evaluation of the School Health activity, the focus of the intervention underwent some changes in order to align with PEPFAR priorities. We were informed that the intervention would target large schools (with over 150 students) in high HIV prevalence districts; this brought into question the external validity of the impact evaluation and the ability to include non-intervention districts with similar characteristics to treatment districts in the design. However, these new criteria do not seem to have affected the actual selection of districts and we will proceed with the original evaluation design. However the number of treatment schools increased. The IP went ahead with the selection of schools for treatment and control before NORC could approve the selection. As a consequence no replacements for control schools were selected.

This can result in a smaller sample than needed. The Evaluation Expert discussed this issue with the IP and USAID.